

Peer-review report of

Tay, L. Q., Hurlstone, M., Jiang, Y., Platow, M. J., Kurz, T., & Ecker, U. K. H. (2024) Causal inference in misinformation and conspiracy research interventions. *advances.in/psychology*, 2, e69941. <https://doi.org/10.56296/aip00023>

Round 1

Dear Authors,

Thank you for submitting your important perspective paper to our special issue in *advances.in/psychology*. I was fortunate to receive evaluations from two long-standing experts in relevant fields. Besides others, one has expertise in misinformation research and the other in causal inference. I thank both for their thorough and constructive reviews.

We all find many strengths in the present paper. It is well-written and argued, focusing on an important issue. Its forward-looking perspective has the potential to provide significant guidance to the field. However, we believe that some of this potential is currently unrealized. I therefore ask you to submit a revised version along with a revision letter in which you respond to each of our comments. In addition to the reviewers' comments, I will provide some points below. Please note that some of these may partly overlap with those of the reviewers.

As Reviewer 2, I stumbled a bit over the section on p. 5, where you outline the fundamental problem of causal inference. Unless the emphasis is strictly on testing two conditions "at the exact same time," multi-trial within-subjects or cross-over designs (I am sure there are many others) seem quite well-suited to experimentally test for individual change. Please elaborate on this point.

Table 1 is great, but I would like to see it extended. First, it would be helpful to add a column explaining the rationale and logic behind each method. Additionally, columns for assumptions, strengths, and weaknesses would be valuable additions. I also think you should present the "standard" approach in the field in the first one or two rows. For space issues, the table could be presented in a horizontal format to incorporate these additions.

Reviewer 1 asks for a figure of the causal effects we may be interested in, which I think is a good idea. Perspective papers often benefit enormously from figures that pedagogically drive home the authors' messages. Thus, I would also strongly

encourage you to create separate figures demonstrating each of the three approaches to testing for causality using the examples presented in the text.

As Reviewer 2, I found the discussion of definitions of misinformation and conspiracy theories somewhat detached from the general arguments presented. I suggest you condense this part and move it to the introduction to briefly set the stage for how you treat these terms.

Finally, a minor issue: Please ensure to include DOIs in URL format for all relevant references (primarily articles), in line with APA7.

Thank you again for submitting this interesting piece. As stated, it is well positioned to make an important contribution to a field that has relied (maybe too much) on a very homogeneous set of methods. I hope that you find the suggested changes meaningful, and I look forward to receiving a revised version within a month.

Best,

Jonas R. Kunst

Editor-in-Chief

Reviewer 1:

This Perspective argues that misinformation and conspiracy research has limited itself to causal claims based on experimental studies and correlational claims, neglecting alternative ways to draw causal inferences. It then provides a brief, non-technical introduction to three types of designs that may be used (instrumental variables, regression discontinuity, difference in difference), drawing on examples from this particular subfield.

I am no misinformation and conspiracy researcher and thus cannot assess whether it is indeed true that natural experiments have been neglected in that field of research, but it seems plausible to me (given that psychologists in general are not very familiar with these designs). I have, however, written extensively on causal inference, in a similar spirit than the authors of the present work. My review thus focuses on matters of causal inference, as well as how to render the manuscript most useful to applied researchers. In general, I think writing causal inference primers for particular audiences is a worthwhile endeavor to improve research practices, and so I do believe that this could be a valuable contribution the field.

Overall, the manuscript is well-written and easy to follow. While technically, it remains on a fairly abstract level, it seems accurate. I would suggest to more clearly introduce what is meant by “correlational design,” and explicitly introduce the term natural experiment as a contrast. I would also suggest to more explicitly refer to papers in which readers can learn more, so that they are not “left hanging” to fill in the gaps themselves. For details on these points as well as additional remarks, please refer to my list of comments below.

Kind regards,

Julia Rohrer

(I sign all my reviews)

Major remarks:

p. 4, “Here, an outcome Y can be said to be caused by event X , if, absent event X , outcome Y would not have occurred.”: This is somewhat confusing – it may either restrict reasoning to the case that outcome Y is binary (0: no occurrence, 1: occurrence), or it may restrict it to cases where X is a necessary cause of Y . Neither restriction is desirable here, so I think a phrasing that works for continuous outcomes is preferable. You could just say something like “Here, an outcome Y can be said to be causally affected by event X , if, absent event X , outcome Y would have taken on a different value.”

p. 5, “In addition, given that what counts as misinformation can often be difficult to define or assess and is instead context-dependent (e.g., assessment might change in light of new evidence), the position that misinformation has no or minimal impact on beliefs and behaviors also implies that information in general has no or minimal impact on beliefs and behaviors...”: This whole part seems a bit of an unnecessary detour, and it may be potentially confusing. I also don’t find the reasoning here particularly convincing: “misinformation is ill-defined and so saying that misinformation has no effects implies that information never has effects.” This matter also isn’t really connected to the question of the definition of causal effects. If you somehow want to connect the definition of causal effects with the claim that it is implausible to claim the misinformation has no effects, I guess you could latch onto the individual-level definition of causal effect: it seems plausible that at least some people do use the misinformation to guide their belief and behaviors, and so even if effects are minimal on average, there would still be causal effects. But I’m not sure this would serve the purpose of the manuscript, I’d just leave the discussion out entirely.

p. 5, “For instance, the adoption of a counterfactual definition means that causal effects at the individual level are impossible to ascertain...”: I don’t think this follows

in any meaningful manner. The counterfactual definition highlights the fundamental problem of causal inference, which is that we cannot observe both potential outcomes at the same time and need to find some workaround. Targeting average effects in randomized experiments is one work-around; but one could also come up with a somewhat plausible identification strategy to ascertain individual level effects for at least some interventions, under the assumption that individuals over time can be used as their own controls. For what it's worth, I think it's always a good idea to introduce people to the fundamental problem of causal inference; I just wouldn't imply that adopting counterfactual notions of causality suddenly means that some effects become impossible to ascertain (it's more like other notions of causality may not even imply the existence of individual-level effects).

p. 8, "Additional Approaches to Causal Inference": Here, it became clear to me that you want introduce readers to natural experiments. That is a good idea, but I think it should be reflected in the abstract (which currently sounds quite generic). Also, you refer to correlational designs without ever making clear what you mean, and the list "mediation" as an approach to which you provide an alternative—this seems rather confusing (mediation is about targeting a specific estimand which a priori does not really imply anything about designs or strategies for causal identification). I think the manuscript would profit if upfront you clearly explain that you slot approaches into three categories—randomized experiments, natural experiments, and "observational studies" by which you seem to mean: non-experimental studies that seek to establish causal identification via third variable adjustment. The first one has limited applicability, the last one may lead to unjustified causal conclusions (and confused analysis goals). The one in the middle still requires assumptions, but it's a nice addition to the causal inference toolbox.

Section on the various designs: In general, these included technical detail – this may have to do with the particular article format and is not a problem per se. However, I think you should provide pointers for readers so that they can build a systematic understanding of the various approaches, rather than trying to piece things together from applied papers. This could be as simple as adding a couple of sentences along the lines of: "A more thorough introduction to this approach can be found in XXX...."

We have recently published a primer to natural experiments that may be helpful given that it has been written for psychologists in particular (Grosz et al., 2024, <https://journals.sagepub.com/doi/10.1177/25152459231218610>). You don't have to cite our paper though, there are plenty of alternative introductions that may work (and should be cited in our paper, in case you need any leads). We don't really talk about difference-in-difference, so you may want to search for a non-technical introduction to it that is still more comprehensive than your short introduction.

Minor remark:

p. 4, “while others have argued that misinformation has the potential to negatively impact

(i.e., a causal claim) societal outcomes”: I think the parentheses don’t quite work here because you couldn’t insert what follows after the “i.e.” in the main sentence. An easy fix that makes this easier to read would be: “has the potential to negatively impact (i.e., to causally affect) societal outcomes.”

Reviewer 2:

Thank you for the opportunity to review this paper. Overall I am happy with it and think it offers valuable contributions to the field, particularly in its call for a more robust set of methodologies. The paper is well-written and clear in its intentions, scope and outline. I am happy to recommend this Perspective for publication, subject to several relatively minor improvements, detailed below.

Throughout: it appears to me that the authors are calling for a methodological integration of the various misinformation-adjacent fields; this is already being done to a degree (the authors already mention several examples but I’d also mention the work by Allcott and Gentzkow from economics, the Alexander Bor/Michael Bang Pedersen and Badrinathan work from political science, and so on); however, perhaps the authors could be more explicit in their argument, which appears to be that psychology is taking up too much breathing room within this field? (I would agree with this personally but the authors might not want to say this outright).

page 5, “Thus, on this view, the argument that misinformation is a symptom rather than a cause, if interpreted strictly, translates to the claim that nothing in our world would change if all misinformation were to disappear” -> this is indeed a very strict interpretation, and I wonder if it holds; for instance, one might say that losing one’s hair is a symptom rather than a cause of old age; but if hair loss were to disappear, then ageing would still work the same but one of its symptoms will have been alleviated. I’d appreciate the authors’ thoughts on this.

Page 5, paragraph starting with “nonetheless”: I wonder if ‘self-report measures’ is the most precise term; a *huge* chunk of the misinformation research (and to a slightly lesser extent also the conspiracy research) relies on item rating tasks that are skills-based (e.g., administering a set of true and false headlines). This is different from self-report scales which (unless I’m confused about the terminology) assess what an individual believes about themselves (e.g., how much misinformation they see online).

Page 6: I appreciate the discussion of causal inference, but it’d be good to see a bit more discussion before this section on what kinds of causal effects we’re interested in within this space. This is not a trivial discussion; perhaps the authors could add a figure explaining their model of causal effects (following Van Bavel et al., 2021 for example)? E.g., misinformation exposure (sometimes) causes belief which (sometimes) causes sharing; and interventions (sometimes) cause reduced belief which in turn reduces sharing (and so on). This is a bad model so please don’t copy

it directly, but it might be useful to contextualise which causal effects are worth trying to measure better and which are not.

Page 8, “Lin and colleagues argued that triangulation of methods (e.g., from self-report to behavioral and physiological) and theories (e.g., from psychology to economics and political science) can be one way of addressing this problem (see also Haslam et al., 2020).” -> This is interesting but it would be good to see a bit more detail, as this is quite high level.

Round 2

Dear Authors,

Thank you for submitting your greatly improved manuscript. Both reviewers were satisfied with the revisions, and I believe we are very close to acceptance. In addition to the two minor points highlighted by Reviewer 2, I would like you to address a few additional minor points:

- Abstract: The abstract can still be more focused on the contributions of your paper, specifically the three methods you propose to complement the current approach. Currently, most of the abstract describes the standard approach and remains vague on the alternative methods that are underutilized. I suggest condensing the sections that focus on existing methods while emphasizing your key points related to them, and then clearly highlighting the suggestions and content inherent to your perspectives paper. Please feel free to reach out if you would like to discuss possible abstract versions.
- Figures: The figures are excellent. However, please ensure they are added to the Word document as lossless SVG files to maintain high resolution. If created in R, exporting them in this format should be straightforward.
- Typos: There are some typos in the current manuscript. For example, in Table 1, there are missing words and minor typos in the second column. Additionally, there is a typo in one of the graphical objects of Figure 1 and in the figure note.
- References: There are some inconsistencies and missing references. Please address all issues highlighted in the citation report, which you can find here <https://drive.google.com/file/d/1CCDYXMJMFDhJpSNgBHPC1pIKkleAv8OD/view?usp=sharing>

Thank you for your continued efforts to make this paper an even more compelling contribution to the field. I believe these remaining minor issues will not take much of

your time and would appreciate receiving a revised version within two weeks, if possible.

Best regards,

Jonas R. Kunst

Reviewer 1:

Thank you for this revised version, which is much improved. I'm happy with the revisions by the authors and appreciate especially the changes based on Reviewer 2's comments. I'm happy to accept this manuscript for publication.

Reviewer 2:

I would like to thank the authors for their responsiveness to my previous comments. I re-read the whole article and really enjoyed it; the newly added figures and the expanded table will make it a lot easier for readers to follow along. The explanations of the IV approach, regression discontinuity, and difference-in-difference are crisp and easy to follow without sacrificing technical accuracy. I only have two very minor phrasing suggestions below, but they are purely cosmetic.

Kind regards,

Julia Rohrer

(I sign all my reviews)

Line 206: „This would allow researchers to test for causal effects as applied to the subset of participants that respond to the instrument; formally, this is known as the complier average causal effect (vs. intention-to-treat analysis, as is common in psychology, which would estimate the average causal effect of the encouragement)“ – I think this sentence could be slightly easier to parse with some light rephrasing so that the “vs.” is followed by something that directly contrasts “complier average causal effect,” so that it is clear what is being contrasted. For example, this could be “complier average causal effect (vs. the average causal effect of the encouragement which would be estimated with an intention-to-treat analysis, the standard in psychology).” Alternatively, you could also consider removing the parentheses and spell it all out in a full sentence, as the contrast may be important enough to highlight it “in the main text.”

Line 262: “Using time as a running variable and the publication of the 1998 Wakefield et al. study falsely linking the MMR vaccine to autism as the cut-off, researchers have used a regression-discontinuity design to examine the causal effect of the fraudulent study on subsequent vaccine skepticism” – I found this grammatical structure somewhat hard to parse, here parentheses or some slight reconstruction could be helpful. For example: “the publication of the 1998 Wakefield et al. study (which falsely linked the MMR vaccine to autism) as the cut-off”